Migration Letters
Volume: 21, No: S7 (2024), pp. 699-716
ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)
www.migrationletters.com

Exponential Life Time Binary Search Tree

Albandari Mohammed Ali Al Badi Aldawsari*

Abstract
The random tree theoryis useful in investigating the equivalent conductance of a random conductance network. Statistical considerations are introduced in network graph theory by replacing the exact average value of trees of the network by the average product of the conductance of a random sample of N-1 branches drawn without replacement from the population of B branches. The case of a binary distribution of conductance has treated and it was found that random trees theory is consistent with effective medium theory applied to the same case. It is supposed that these periods of times are independent identically distributed with exponential distribution with parameter 1. The essential focus is on the successful and unsuccessful searching. For the introduced random variables the mean is given, variance and asymptotic distribution. Law of large numbers is also established.

***Keywords**: Random Tree, Binary Search Tree, Exponential Distribution.*

1. Introduction

Each model carries a coherent set of algebraic and analytic techniques, which we illustrated by reviewing a few c[1]haracteristic examples. In mathematics and computer science, a random tree is a tree or arborescence that is formed by a stochastic process. The theory of records in sequences of independent identically distributed random variables leads to simple proofs of various properties of random trees, including the limit law for the depth of the last node of randomly ordered trees, random union-find trees, and random binary search trees. we will presented classes of random tree models that occur in the average case analysis of a variety of computer algorithms, including symbolic manipulation algorithms, comparison based on searching and sorting digital retrieval techniques, systems and communication protocols. Each model carries a coherent set of algebraic and analytic techniques, which we illustrated by reviewing a few characteristic examples. In mathematics and computer science, a random tree is a tree or arborescence that is formed by a stochastic process. The theory of records in sequences of independent identically distributed random variables leads to simple proofs of various properties of random trees, including the limit law for the depth of the last node of randomly ordered trees, random union-find trees, and random binary search trees.

In This paper, we consider a binary search tree and we study some of their practical aspect. The interest concerns here the natural growth of binary search trees and it is proposed to give some new analytical results under some realistic hypotheses. The hypotheses tell that every external node have gives two new nodes after a random period of time. It is supposed that these periods

* Department of Mathematics, Faculty of Science and Human studies, Al- Kharj, Prince Sattam Bin Abdul Aziz University.

of times are independent identically distributed with exponential distribution with parameter 1. The essential focus is on the successful and unsuccessful searching. For the introduced random variables the mean is given, variance and asymptotic distribution. Law of large numbers is established.

2. Unsuccessful searching

Notations: For a binary search tree T_n with size n we denote by

(a) $\tilde{T}_n$ the number of comparisons consumed by an unsuccessful search in T_n to insert a new key.

(b) X_{nk} the number of external nodes at level k of T_n.

The model: In our model we suppose that to go from a node x to any neighbor node y we consume a random time having Exponential distribution with parameter 1. We suppose also that all these Exponential random variables are independent . This is our extension comparing with M. Hosam 1992.

Remark 1.

(rm1) $T_1, T_2, T_3, T_4, \cdots, T_n$ are iid random variables with exponential distribution with parameter 1.

(rm2) Let D_k the depth of x_k in T_n

(rm3) for our model $\tilde{T}_k$ can be written as

$$\tilde{T}_k = T_1 + T_2 + \cdots + T_{D_k}$$

Proposition 1. The random variable $\tilde{T}_n$ is a continuous random variable with probability density function $\tilde{f}_n$ given by

$$\tilde{f}_n(x) = \sum_{k=1}^{n-1} \frac{2_k}{(n+1)!} \binom{n}{k} \frac{x^{k-1}}{T(k)} \exp(-k), \text{ for } x \geq 0.$$

Proof: We have $\tilde{T}_n = T_1, T_2, T_3, T_4, \ldots\ldots\ldots, T_{D_n}$. Let $\tilde{f}_n(x) = \mathbf{P}(\tilde{T}_n \leq x)$ be the cumulative distribution function of $\tilde{T}_n$. We know that $\tilde{f}_n(x) = \frac{\partial}{\partial x}\tilde{F}_n(x)$. We have

$$\tilde{F}_n(x) = \mathbf{P}\left(\sum_{i=1}^{D_n} T_i \leq x\right)$$

$$= \sum_{k=1}^{n-1} \mathbf{P}\left(\left[\sum_{i=1}^{k} T_i \leq x\right] \cap [D_n = k]\right)$$

$$= \sum_{k=1}^{n-1} (D_n = k) \times \mathbf{P}\left(\sum_{i=1}^{k} T_i \leq x\right)$$

The last equality is due to the fact that D_n is independent of the sequence $(T_1, T_2, T_3, T_4, \cdots T_k)$. But using theorem of Lynch (1965) we know that $\mathbf{P}(D_n = k) = \frac{2^k}{(n+1)!}\binom{n}{k}$. On the other hand we know that for all integer k the summation

$T_1 + \cdot \cdot \cdot + T_k$ is distributed like Gamma(k, 1). Then we conclude that

$$\tilde{F}_n(x) = \sum_{k=1}^{n-1} \frac{2k}{(n+1)} \binom{n}{k} \int_0^x \frac{y^{k-1}}{T(k)} \exp(-y)\, dy.$$

and we deduce the value of $\tilde{F}_n(x)$ by differentiating.

Proposition 2. The mean of $\tilde{T}_n$ is given by

$$[\tilde{T}_n] = 2\left[\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n+1}\right]$$

$$= 2[S_{n+1} - 1] \approx 2\text{In}(n+1),$$

Where $S_n = [1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}]$.

Proof : the probability density function $\tilde{f}_n(x)$ can be written as

$$\frac{1}{(n+1)!\,(k-1)!} \sum_{k=1}^{n-1} 2^k x^{k-1} \exp(-x) \binom{n}{k}.$$

Then :

$$E[\tilde{T}_n] = \frac{1}{(n+1)!\,(k-1)!} \sum_{k=1}^{n-1} \int_0^{\infty} 2^k x^k e^{-x} \binom{n}{k}\, dx$$

$$= \frac{1}{(n+1)!\,(k-1)!}\sum_{k=1}^{n-1} 2^k \binom{n}{k} \int_0^{\infty} x^k e^{-x} dx$$

$$= \frac{1}{(n+1)!\,(k-1)!}\sum_{k=1}^{n-1} 2^k \binom{n}{k} T(k+1)$$

$$= \frac{1}{(n+1)!}\sum_{k=1}^{n-1} 2^k k \binom{n}{k} = \frac{2}{(n+1)!}\sum_{k=1}^{n-1} 2^{k-1} k \binom{n}{k}$$

$$= \frac{2}{(n+1)!} f(2)$$

Where $f(2) = \sum_{k=1}^{n-1} 2^{k-1} k \binom{n}{k}$. Put $(x) = \sum_{k=1}^{n-1} x^k \binom{n}{k}$, then

$$g'(x) = \sum_{k=1}^{n-1} x^{k-1} k \binom{n}{k} = f_n(x).$$

On the other hand (x) can be written as

$$(x) = x(x+1)\cdots(x+n-1).$$

By differentiating the function g and replacing x by 2 we conclude that

$$f(2) = (n+1)!\,[\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n+1}].$$

We conclude the result of the proposition.

Proposition 3. The variance of $\tilde{T}_n$ is given by

$$Va[\tilde{T}_n] = 4S_{n+1} - 5,$$

Where $S_n = [1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}]$.

Proof: we have

$Va(\tilde{T}_n) = E(\tilde{T}_n^2) - (E(\tilde{T}_n))^2$

$(E(\tilde{T}_n))^2 = 4\,[\frac{1}{(2)} + \frac{1}{(3)} + \frac{1}{(4)} + \cdots + \frac{1}{(n+1)}]^2$

And

$$((\tilde{T}_n)^2) = \int_0^\infty x^2 f_{T_n}(x)dx = \sum_{k=1}^{n-1} \frac{2^k \binom{n}{k}}{(n+1)!\,(k)} \int_0^\infty x^2 x^{k-1} e^{-x} dx$$

$$= \sum_{k=1}^{n-1} \frac{2^k \binom{n}{k}}{(n+1)!\,(k)} \int_0^\infty x^{k+1} e^{-x} dx = \sum_{k=1}^{n-1} \frac{2^k \binom{n}{k}}{(n+1)!\,T(k)} (k+2)$$

$$= \sum_{k=1}^{n-1} \frac{2^k}{(n+1)!} \binom{n}{k} \frac{(k+1)!}{(k-1)!} = 2 \sum_{k=1}^{n-1} \frac{2^{k-1}}{(n+1)!} (k+1)(k) \binom{n}{k}.$$

Let the function f defined by $(x) = \sum_{k=1}^{n-1} x^{k+1} \binom{n}{k}$, then

$$f'(x) = \sum_{k=1}^{n-1} (k+1)x^k \binom{n}{k} \; and \; f''(x) = \sum_{k=1}^{n-1} (k)(k+1)x^{k-1} \binom{n}{k}$$

and we remark that $E((\tilde{T}_n)^2) = \frac{2}{(n+1)!} f''(2)$.

On the other hand we have

$$f(x) = x \sum_{k=1}^{n-1} x^k \binom{n}{k} \quad ,$$

$$f(x) = x \sum_{k=1}^{n-1} x^k \binom{n}{k}$$

$$(x) = x^2 \mathbf{G}_{k=1}^{n-1} (x+k)$$

$$(x) = x^2 \mathbf{G}_{k=1}^{n-1} (x+k).$$

Then

$$f'(x) = 2x \prod_{k=1}^{n-1}(x+k) + x^2 \sum_{k=1}^{n-1} \prod_{l=1,\neq k}^{n-1} (x+l)$$

$$f''(x) = 2 \prod_{k=1}^{n-1}(x+k) + 2x \sum_{k=1}^{n-1} \prod_{l=1,\neq k}^{n-1} (x+l)$$

$$+2x \sum_{k=1}^{n-1} \prod_{l=1,\neq k}^{n-1} (x+l)$$

$$= x^2 \sum_{k=1}^{n-1} \sum_{l=1,\neq k}^{n-1} \prod_{j=1,j\neq l,j\neq k}^{n-1} (x+j)$$

$$f''(2) = 2 \prod_{k=1}^{n-1}(2+k) + 8 \sum_{k=1}^{n-1} \prod_{l=1,\neq k}^{n-1} (2+l) + 4 \sum_{k=1}^{n-1} \sum_{l=1,\neq k}^{n-1} \prod_{j=1,j\neq l,j\neq k}^{n-1} (2+j)$$

$$= 2 \times [3 \times 4 \times \ldots \times (n+1)]$$

$$+\frac{8}{2}\sum_{k=1}^{n-1} \frac{2 \times 3 \times 4 \times \ldots \times (n+1)}{(2+k)} + \frac{4}{2}\sum_{k=1}^{n-1}\sum_{L=1}^{n-1} \frac{2 \times 3 \times 4 \times \ldots \times (n+1)}{(2+L)(2+k)}$$

$$(n+1)! + 4 \times (n+1)! \sum_{k=1}^{n-1} \frac{1}{2+k} + 2(n+1)! \sum_{k=1}^{n-1} \frac{1}{2+k} \sum_{L=1}^{n-1} \frac{1}{2+L}$$

$$= (n+1)!\,[1 + 4 \sum_{i=3}^{n+1} \frac{1}{i} + 2(\sum_{m=3}^{n+1} \frac{1}{m})(\sum_{p=3}^{n-1} \frac{1}{p})]$$

$$= (n+1)!\,[1 + 4\,(S_{n+1} - 1 - \frac{1}{2}) + 2\,(S_{n+1} - 1 - \frac{1}{2})\,(S_{n+1} - 1 - \frac{1}{2})]$$

$$= (n+1)!\,[1 + 4S_{n+1} - 4 - 2 + 2(S_{n+1} - \frac{3}{2})^2\,]$$

$$E((\tilde{T}_n)^2) = \frac{2}{n+1!} f''(2)$$

$$E((\tilde{T}_n)^2) = \frac{2}{(n+1)!} \times (n+1)!\,[1 + 4S_{n+1} - 6 + 2(S_{n+1} - \frac{3}{2})^2]$$

Then

$$Va(\tilde{T}_n) = E(\tilde{T}_n^2) - (E(\tilde{T}_n))^2$$

$$= -10 + 8S_{n+1} + 4(S_{n+1} - \frac{3}{2})^2 - 4(S_{n+1} - 1)^2$$

$$= -10 + 8S_{n+1} + 4\left[S_{n+1}^2 + \frac{9}{4} - 3S_{n+1} - S_{n+1}^2 - 1 + 2S_{n+1}\right]$$

$$= -10 + 8S_{n+1} + 9 - 4S_{n+1} - 4$$

$$= -5 + 4S_{n+1} = 4S_{n+1} - 5$$

$$Va(\tilde{T}_n) = 4S_{n+1} - 5$$

Theorem 1. The random variable $\frac{\tilde{T}_n - D}{\sqrt{2D_n}}$ converges in distribution to a standard normal variable with mean 0 and variance 1.

Proof:

We know that

(a) $\frac{D_n}{2\,In(n)} \xrightarrow[n \to +\infty]{P} 1,$

(b) and if $T_1, T_2, T_3, T_4, \cdots$ are iid with distribution Exp(1), then $\frac{W_n - n}{\sqrt{n}}$ converges in distribution to a standard normal variable with mean 0 and variance 1,

where $W_n = \sum_{k=1}^{n} T_k$.

Then

$$\frac{\tilde{T} - D_n}{\sqrt{D_n}} = \frac{W_{D_n} - 2\,In(n)}{\sqrt{D_n}} + \frac{2\,In(n) - D_n}{\sqrt{D_n}}$$

At first we have (Brown and Shubert, 1984)

$$\frac{D_n - 2\,I(n)}{\sqrt{D_n}} = \frac{D_n - 2\,In(n)}{\sqrt{2I(n)}} \times \sqrt{\frac{2In(n)}{D_n}} \to N_1(0,1) \times 1$$

the other have we write :

$$\frac{W_{D_n} - 2I(n)}{\sqrt{D_n}} = \frac{W_2 I(n) - 2In(n)}{\sqrt{D_n}} + \frac{W_{D_n} - W_2 I(n)}{\sqrt{D_n}}$$

we have
(i) using(b) $\tilde{n} = In(n)$ we can deduce

$$\frac{W_{2I(n)} - 2In(n)}{\sqrt{D_n}} = \frac{W_{2I(n)}}{\sqrt{2I(n)}} \times \sqrt{\frac{2I(n)}{D_n}} \to N_2(0,1) \times 1$$

(ii) For the second

$$\left|\frac{W_{Dn} - W_{2I(n)}}{\sqrt{D_n}}\right| = \left|\frac{\sum_{k=1}^{D_n} T_k - \sum_{k=1}^{2I(n)} T_k}{\sqrt{D_n}}\right|$$

$$\leq \left|\frac{\sum_{k=D_n}^{2I(n)} T_k}{\sqrt{D_n}}\right|$$

$$\leq \left|\frac{2I(n) - D_n \lceil max D_n \leq k \leq 2In(n) \rceil T_k}{\sqrt{D_n}}\right|$$

$$\leq \left|\frac{2I(n) \left|1 - \frac{D_n}{2I(n)}\right| max D_n \leq k \leq 2I(n) |T_n|}{\sqrt{D_n}}\right|$$

We conclude that:

$$\frac{\tilde{T}_n - D_n}{\sqrt{D_n}} \to N_1(0,1) + N_2(0,1)$$

We have $N_1(0,1)$ and $N_2(0,1)$ are independent because T$_1$, T$_2$, T$_3$, T$_4$, · · · is sequence and independent of D_n. But $N_1(0,1) + N_2(0,1) \overset{\mathcal{D}}{=} N(0,2)$.

Finally $\frac{\tilde{T}_n - D_n}{\sqrt{2D}} \overset{\mathcal{D}}{=} (0,1)$.

3. Successful Searching

Definition 1. Let T_n be a Binary Search Tree withe size n (having n internal nodes so $n + 1$ external nodes) . We denote by $X_{n,}$ the random number of external nodes at level k. The random variable $M_{n,}$ denotes the number of internal node at level k.

Example 1. For the following tree we have

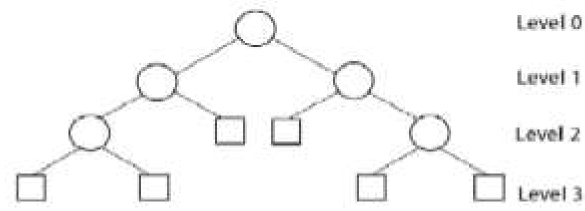

Figure 3.1: A complete binary tree

$$X_{6,0} = 0, \qquad X_{6,1} = 0,$$

$$X_{6,2} = 2, \qquad X_{6,3} = 4,$$

And

$$M_{5,0} = 1, \quad M_{5,1} = 2,$$

$$M_{5,2} = 2, \quad M_{5,3} = 0.$$

Proposition 4. The number of internal nodes at level k satisfies the following recursion : For all k = 0, 1, 2, · · · , $n - 1$ we have

$$M_{n,k} = \sum_{j=k+1}^{n} \frac{X_{n,}}{2^{j-k}}.$$

Proof: By induction,for $k = n$ it is clear that $M_{n,n} = \sum_{j=n+1}^{n} \frac{X_{n,}}{2^{j-n}} = 0.$

For $k \in \{0,\cdots,n-1\}$ put $N_{n,k} = \sum_{j=k+1}^{n} \frac{X_{n,j}}{2^{j-k}}.$

We prove that , $N_{n,k} = M_{n,k}$, and

$$N_{n,+1} + X_{n,k+1} = \sum_{j=k+2}^{n} \frac{X_{n,j}}{2^{j-(1+k)}} + X_{n,k+1}$$

$$= \sum_{j=k+2}^{n} \frac{X_{n,j}}{2^{j-1-k}} + X_{n,+1}$$

$$= \sum_{l=k+1}^{n} \frac{X_{n,l+1}}{2^{l-k}} + \frac{X_{n,k+1}}{2^{k-k}} = \sum_{l=k}^{n} \frac{X_{n,l+1}}{2^{l-k}}$$

$$= \sum_{j=k+1}^{n} \frac{X_{n,j}}{2^{j-1-k}} = 2 \sum_{j=k+1}^{n} \frac{X_{n,j}}{2^{j-k}} = 2M_{n,.}$$

So $N_{n,}$ satisfies (3.1) with $N_{n,n} = 0$. Then $N_{n,} = M_{n,k} \forall k \in \{0, \dots, n-1\}$.

Theorem 2. (Lynch (1965)) We have

$$E[M_{n,k}] = \sum_{j=k+1}^{n} \frac{E[X_{n,j}]}{2^{j-k}} = \frac{2^k}{n!} \sum_{j=k+1}^{n} \binom{n}{j}$$

We now turn to study the depth of an internal node. denote by $\tilde{S}_n$ the depth of an internal node chosen at random. $\tilde{S}$is the time required to reach an internal node chosen at random. Then it is clear that for our model, we have:

$\tilde{Q} = T_1 + T_2 + T_3 + T_4 + \cdots + T_{S_n}$, where, $T_1, T_2, T_3, T_4, \dots, T_n, \dots$, are iid with exponential distribution with parameter 1 and $\tilde{S}_n$ is the depth of an internal node chosen at random as in the book by Mahmoud, H. (1992).

Lemma 1. The process $M_{n,}, 0 \le k \le n$ satisfies the following recursion :

$$2M_{n,} = M_{n,k+1} + X_{n,k+1}, \qquad k \le n-1,$$

with the condition

$$M_{n,} = 0, \qquad M_{n,0} = 1.$$

Let $\tilde{Q}$be the time consumed to successful searching at random an internal with depth S_n node . It is easily to see that $\tilde{Q}= T_1 + T_2 + T_3 + \cdots + T_{S_n}$
Where , as in the last section , $T_1, T_2, T_3, \dots$ are independent identically distributed random variable with exponential distribution with parameter 1.as in the lest section we propose to compute the mean , the variance and the probability density function of $\tilde{Q}$.

Theorem 3. The probability density function of $\tilde{Q}$is given by

$$f_{\tilde{Q}_n}(x) = \frac{1}{n}\sum_{k=1}^{n-1}\frac{(2x)^{-1}}{(k-1)!}e^{-x}\sum_{j=k}^{n}\binom{n}{j}, \forall x \geq 0.$$

Proof:

$$f_{\tilde{Q}}(\boldsymbol{x}) = F'_{\tilde{Q}_n}(x) = \boldsymbol{P}(\tilde{Q}_n \leq x)$$

$$F_{\tilde{Q}_n}(x) = \boldsymbol{P}(\sum_{i=1}^{S_n}\tilde{T}_i \leq x)$$

$$= \sum_{k=1}^{n-1}\boldsymbol{P}((\sum_{i=1}^{k}\tilde{T}_i \leq x) \cap (S_n = k))$$

$$F_{\tilde{Q}_n}(x) = \sum_{k=1}^{n-1}\boldsymbol{P}(S_n = k)\boldsymbol{P}(\sum_{i=1}^{k}\tilde{Q} \leq x)$$

According to (Brown and Shubert 1984) we have

$$\boldsymbol{P}(S_n = k) = \frac{(2)^{k-1}}{n(n!)}\sum_{i=k}^{n}\binom{n}{j}$$

$$(\sum_{i=1}^{k}\tilde{Q} \leq x) = \boldsymbol{P}(X \leq x) \qquad X \sim T_{(n,1)}$$

$$\boldsymbol{P}(X \leq x) = \int_0^x \frac{y^{k-1}}{(k)}e^{-y}dy$$

$$F_{\tilde{Q}_n}(x) = \sum_{k=1}^{n-1}\frac{2^{k-1}}{n(n!)}\sum_{j=k}^{n}\binom{n}{j}\int_0^x \frac{y^{k-1}}{T(k)}e^{-y}dy$$

therefore

$$f_{\tilde{Q}_n}(x) = \sum_{k=1}^{n-1}\frac{2^{k-1}}{n(n!)}\sum_{j=k}^{n}\binom{n}{j}\frac{x^{k-1}}{T(k)}e^{-x} \qquad x \geq 0.$$

to obtain the expected value of $\tilde{Q}_n(x)$ we can use the probability function , using equation

$$E[X_{n,j}] = \frac{2^j \binom{n}{j}}{n!}$$

Using the last equation we are able to compute$[\tilde{Q}]$:

Lemma 2. The expected value of $\tilde{Q}$is given by :

$$E[\tilde{Q}_n] = E[S_n] = 2\,[1 + \frac{1}{n}]\,H_n - 3,$$

Where

$$H_n = (1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}).$$

Proof: we have

$$\tilde{Q} = T_1 + T_2 + T_3 + \cdots + T_{S_n}$$

$$= \sum_{k=1}^{S_n} T_k = \sum_{l=1}^{n-1} \sum_{k=1}^{l} T_k 1_{S_n = l}.$$

The sequence $(T_1, T_2, \ldots)$ is independent of S_n, then :

$$E[\tilde{Q}] = \sum_{l=1}^{n-1} \sum_{k=1}^{l} E[T_k] \boldsymbol{P}(S_n = l) = \sum_{l=1}^{n-1} \boldsymbol{P}(S_n = l).$$

We conclude the proof using the results in the book of Hosam M. (1992)

Remark 4. The random variables $Q_1, Q_2, \ldots, Q_k$ are independent and identically distributed with exponential distribution with parameter 1. A well known result says that

$Q_1 + Q_2 + \cdots + Q_k$ have as distribution $(k, 1)$,
where the probability density function is given by$f_{Q1+Q2+\cdots+Qk}(x) = e^{-x} \frac{x^{k-1}}{T(k)}$ This gives the probability density function of $\tilde{Q}$:

$$f_{\tilde{Q}}(x) = \sum_{k=1}^{n-1} \frac{2^{k-1}}{n(n!)} \sum_{j=k}^{n} \binom{n}{j} \frac{x^{k-1}}{T(k)} e^{-x} \qquad x \geq 0.$$

Using this probability density function we can write

$$E(\tilde{Q}_a) = \sum_{k=1}^{n-1} \frac{2^{k-1}}{n(n!)} \sum_{j=k}^{n} \binom{n}{j} \int_0^x \frac{x^k}{T(k)} e^{-x} dx = \sum_{k=1}^{n-1} \frac{2^{k-1}}{n(n!)} \sum_{j=k}^{n} \binom{n}{j} \frac{(k+1)}{T(k)}$$

$$= \sum_{k=1}^{n-1} \frac{2^{k-1}}{n(n!)} \sum_{j=k}^{n} \binom{n}{j} (k+1) = \frac{k+1}{n(n!)} \sum_{k=1}^{n-1} 2^{k-1} \sum_{j=k}^{n} \binom{n}{j}.$$

$$= \frac{(k+1)}{n(n!)} (x) \; where \; f(x) = \sum_{k=1}^{n-1} 2^{k-1} \sum_{j=k}^{n} \binom{n}{j}.$$

Theorem 4. The variance of $\tilde{Q}$ is given by :

$$Va[\tilde{Q}]_n = (4 + \frac{12}{n})_n - 4\left(1 + \frac{1}{n}\right)\left[-\frac{H_n^2}{n} + H2_n\right].$$

Proof:We have

$$\tilde{Q} = \sum_{k=1}^{S_n} T_k, (\tilde{Q}_a)^2 = (\sum_{k=1}^{S_n} T_k)^2 = (\tilde{Q}_a)^2 = \sum_{l=1}^{n-1} 1_{S_n = l} (\sum_{k=1}^{l} T_k)^2$$

Because $Va(\tilde{Q}_a) = \boldsymbol{E}[(\tilde{Q}_a)^2] - [\boldsymbol{E}(\tilde{Q}_a)]^2$ we have:

$$[(\tilde{Q}_a)^2] = \sum_{l=1}^{n-1} \boldsymbol{E}([1_{S_n = l}](\sum_{k=1}^{l} T_k)^2)$$

$$= \sum_{l=1}^{n-1} [1_{S_n = l}](\boldsymbol{E} \sum_{k=1}^{l} T_k)^2)$$

$$= \sum_{l=1}^{n-1} (S_n = l) \boldsymbol{E}(\sum_{1 \le k_{1,2} \le l} T_{k1} T_{k2})$$

$$= \sum_{l=1}^{n-1} (S_n = l)(\sum_{1 \le k1,2 \le l} E[T_{k1} T_{k2}])$$

$$= \sum_{l=1}^{n-1} (S_n = l)[\sum_{k=1}^{l} E[T^2_{k1}] + \sum_{1 \le k1 \ne k2 \le l} E[T_{k1}]E[T_{k2}]]$$

$$= \sum_{l=1}^{n-1} (S_n = l)[2l + \sum_{1\leq k1 \neq k2 \leq l} 1]$$

$$= \sum_{l=1}^{n-1} (S_n = l)[2l + l^2 - l]$$

$$= \sum_{l=1}^{n-1} (l^2 + l)\boldsymbol{P}(S_n = l)$$

$$= \sum_{l=1}^{n-1} l^2(S_n = l) + \sum_{l=1}^{n-1} l\boldsymbol{P}(S_n = l)$$

$[(\tilde{Q}_n)^2] = \boldsymbol{E}[S_n]^2 + E[S_n]$.

Finally

$$Va[\tilde{Q}_n] = \boldsymbol{E}[(\tilde{Q}_n)^2] - \boldsymbol{E}[\tilde{Q}_n]^2$$

$$= [S_n^2] + \boldsymbol{E}[S_n] - \boldsymbol{E}[S_n]^2$$

$$= [S_n^2] - \boldsymbol{E}[S_n]^2 + \boldsymbol{E}[S_n]$$

$$= (4 + \frac{12}{n}H_n) - 4(1 + \frac{1}{n})[\frac{2}{n} + H_n^2].$$

Note:

$$[\tilde{Q}_n] \simeq 2In(n), Var[\tilde{Q}_n] \simeq 4In(n),$$

We deduce

$$\frac{\tilde{Q}_n}{In(n)} \underset{n \to +\infty}{\overset{\boldsymbol{P}}{\to}} 2.$$

4. Internal and external path lengths

Let T_n be a binary tree with n internal nodes . for all x an internal node denote by L_x be the consumed time to search x . for the particular case . if $x = \emptyset$: the root, by convention we have$l_\emptyset = 0$. the internal path length of T_n . the quantity denoted by I_n defined then as :

$$I_n = \sum_{x \in T_n, x internal\, node} l_x$$

The process I_n is used to give a measure representing the amount of time consumed to search all the internal nodes of T_n. for T_n We know that, we have

$n+1$ leaves (external nodes). let for y external node of T_n , x_y be the unsuccessful searching of y . Then the sum of all unsuccessful searching of all the external nodes of T_n is denoted by J_n defined as :

$$J_n = \sum_{y \in \partial \tilde{T}_n} x_y$$

Where $\partial \tilde{T}_n$: is the set of leaves of T .

The external path length $\tilde{J}_n$ serves as a measure of all unsuccessful searching of all the external nodes of T_n.

For our model: $\tilde{I}_{n+1} = \tilde{I}_n + \sum_{j=1}^{sk} T_k$

$$\tilde{J}_{n+1} = \tilde{J}_n - \sum_{j=1}^{s_k} T_j + 2\sum_{j=1}^{s_k} T_j + T^{(1)} + T^{(2)}$$

Conclusion : $\tilde{I}_{n+1} = \tilde{I}_n + \sum_{j=1}^{sk} T_j$

$$\tilde{J}_{n+1} = \tilde{J}_n + \sum_{j=1}^{s_k} T_j + T^{(1)} + T^{(2)}.$$

Our aims is to obtain the relation between

$$\tilde{I}_{n+1} \text{ and } \tilde{J}_{n+1}$$

Theorem 5.

$$\tilde{J}_n = \tilde{I}_n + \sum_{j=1}^{2n} T_j$$

Proof: The proof is by induction

• for $n = 1$, the only tree with one internal node is the following:

$$I_1 = 0 \quad x_1 = x_1 + x_2 = 2$$

$$I_1 = 0 + \sum_{k=1}^{2} T^k = T^1 + T^2$$

• suppose the relation is true for all binary tree with size n. we prove is also true for all binary tree with size $n + 1$. We have

$$\tilde{J}_{n+1} = \tilde{J}_n + \sum_{j=1}^{s_k} T_j + T^{(1)} + T^{(2)}$$

$$= \tilde{I}_n + \sum_{j=1}^{2n} T_j + \sum_{j=1}^{s_k} T_j + T^{(1)} + T^{(2)}$$

$$= \tilde{I}_{n+1} + \sum_{j=1}^{2n} T_j + T^{(1)} + T^{(2)}$$

$$\tilde{I}_{n+1} + \sum_{j=1}^{2(n+1)} T_j$$

Theorem 6. We have

$$[\tilde{S}_n] = \frac{1}{n}\mathbf{E}[\tilde{I}_n] + 1.$$

Proof:

$$E[\tilde{S}_n / \tilde{T}_0, \tilde{T}_1, \ldots, \tilde{T}_{n-1}] = \frac{(L_0 + T_0)}{n} + \frac{(L_1 + T_1)}{n} + \cdots + \frac{(L_{n-1} + T_{n-1})}{n}$$

$$= \frac{1}{n}\sum_{k=0}^{n-1} L_k + \frac{1}{n}\sum_{k=0}^{n-1} T_k.$$

Taking expectations of (3.2)

$$[\tilde{S}_n] = \frac{1}{n}\mathbf{E}[\sum_{k=0}^{n-1} L_k] + \frac{1}{n}\mathbf{E}[\sum_{k=0}^{n-1} T_k]$$

$$= \frac{1}{n}[\tilde{I}_n] + \frac{1}{n}(n)$$

$$= \frac{1}{n}\mathbf{E}[\tilde{I}_n] + 1$$

$$E[\tilde{J}_n / \tilde{T}_0, \tilde{T}_1, \ldots, \tilde{T}_{n-1}] = \frac{J_1}{n+1} + \frac{J_2}{n+1} + \cdots + \frac{J_{n+1}}{n+1}$$

$$= \frac{1}{n+1} \sum_{i=1}^{n+1} J_n$$

$$= \frac{J_n}{n+1}$$

$$= \frac{(J_n)}{n+1}.$$

Substituting the new expressions for $[J_n]$ and $\mathbf{E}[I_n]$

$$[J_n] = \mathbf{E}[I_n] + \mathbf{E}[\sum_{j=1}^{2n} T_j] = \mathbf{E}[I_n] + 2n.$$

Conclusion:

This paper had studied the random binary search trees but under exponential distribution. It was essentially interested to the time insertion of a given node, the first time to reach some given level, the first time that some given level becomes full: the first time saturation. In the future we emphasize to study the random exponential binary tree EBT introduced by Feng a_nd Mahmoud (2017).

References:

[1] Panholzer, A.; Prodinger, H.: Spanning tree size in random binary search trees. Ann. Appl. Probab. 14(2), 718-733 (2004).

[2] Aguech, R.; Amri, A.; Sulzbach, H.: On weighted depths in random binary search trees.Journal of Theoretical Probability, 1-23 (2017).

[3] Aguech, R.; Lasmar, N.; Mahmoud, H.: Extremal weighted path lengths in random binarysearch trees. Probab. Engrg. Inform. Sci., 21(1), 133-141 (2007).

[4] Brown, G.; shubert, B.: On random binary trees. Mathematics of Operations Research, (9),43-65 (1984).

[5] Billingsley, P.: Convergence of probability measures. Wiley Series in Probability and Statistics:Probability and Statistics. John Wiley and Sons, Inc., New York, 1999.

[6] Devroye, L,: A note on the height of binary search trees. J. Assoc. Comput. Mach., 33(3),489-498 (1986).

[7] Devroye, L.; Neininger, R.: Distances and finger search in random binary search trees. SIAMJ. Comput., 33(3), 647-658 (2004) .

[8] Grübel, R.: On the silhouette of binary search trees. Ann. Appl. Probab., 19(5), 1781-1802 (2009).

[9] Mahmoud, H.: Evolution of random search trees. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley and Sons, Inc., New York (1992).

[10] Mahmoud, H.; Neininger, R.: Distribution of distances in random binary search trees. Ann.Appl. Probab., 13(1), 253-276 (2003).

[11] Mahmoud, H.; Pittel, B.: On the most probable shape of a search tree grown from a randompermutation. SIAM J. Algebraic Discrete Methods, 5(1), 69-81 (1984).

[12] Feng, Y.; Mahmoud, H.: Profile of Random Exponential Binary Trees. Methodol Comput Appl Probab DOI 10.1007/s11009-017-9578, (2017).

[13] Kuba, M.; Panholzer, A.: On weighted path lengths and distances in increasing trees. Probab.Engrg. Inform. Sci., 21(3), 419-433 (2007).

[14] Lynch, W.: More combinatorial problems on certain trees.The Computer Journal, Vol. 7,299-302 (1965).

Acknowledgement: The author is thankful to the Research and Innovation Agency of the Ministry of Education of the Kingdom of Saudi Arabia. Also, he expresses his appreciation to Prince Sattam bin Abdulaziz University. Finally, the author thanks the anonymous referees for their useful and constructive comments that improved this paper.